

THE RELATION OF GENDER, PERSONALITY, AND INTELLIGENCE TO JUDGES' ACCURACY IN JUDGING STRANGERS' PERSONALITY FROM BRIEF VIDEO SEGMENTS

Richard A. Lippa and Joshua K. Dietz

ABSTRACT: Fifty-three men and 56 women viewed brief video segments of 32 male targets and rated them on three personality traits: extraversion, neuroticism, and masculinity-femininity (M-F). Judges were assessed on general intelligence, Big Five traits, and gender-related traits. Two measures of accuracy were computed: 1) consensus accuracy, which measured the correlation between judges' ratings and corresponding ratings made by previous judges, and 2) trait accuracy, which measured the correlation between judges' ratings and targets' assessed personality. There was no gender difference in overall accuracy. However, women showed higher trait accuracy than men in judging neuroticism. Consensus accuracy exceeded trait accuracy, and extraversion and M-F were judged more accurately than neuroticism. M-F judgments showed the highest level of consensus accuracy. Judges' intelligence correlated positively with accuracy. Except for openness, personality traits were generally unrelated to accuracy.

Some of the earliest research on the accuracy of personality judgments focused on characteristics of the "good judge" (see Taft, 1955, for a review). This research came to an abrupt halt in the mid-1950s, in part because of weak and inconsistent results, but more fundamentally because of serious methodological problems in the assessment of accuracy (Davis & Kraus, 1997; Funder, 1995). Cronbach's (1955) much-cited statistical critique showed that, when computed as squared deviations between personality judgments and personality criteria, accuracy measures comprised a number of components (e.g., differential accuracy, stereotype accuracy, and the effects of response sets such as judges' tendencies to use non-extreme vs. extreme responses). After a long period of dormancy, research

Richard A. Lippa and Joshua K. Dietz, Psychology Department, California State University.

Address correspondence to Richard A. Lippa, Psychology Department, California State University, Fullerton, CA 92834.

on the accuracy of personality judgments resumed in the 1980s when researchers developed more sophisticated methods to assess accuracy and thereby overcome the methodological problems that plagued earlier research (Funder, 1995; Funder & West, 1993; Kenny, 1991, 1994; Kenny & Albright, 1987; Park & Judd, 1989). Renewed interest was also fostered by the emergence of the five-factor model of personality, which provided not only a framework for organizing the traits that judges were asked to perceive in others but also for assessing the personality traits of judges.

The latest wave of research has studied a range of factors that influence accuracy, including the kinds of traits being judged, the characteristics of judges, the information judgments are based on, and the nature of the relationship between the judge and the target (Colvin, 1993; Funder, 1995; Borkenau & Liebler, 1993). Recent research on the judgment of strangers in zero-acquaintance situations suggests, for example, that extraversion and sociability are more accurately judged than other Big Five traits (Albright, Kenny, & Malloy, 1988; Borkenau & Liebler, 1993; Funder & Colvin, 1988; Funder & Dobruth, 1987; Kenny, Horner, Kashy, & Chu, 1992; Levesque & Kenny, 1993; Watson, 1989). Not surprisingly, judges' accuracy tends to increase as they have more contact with the people they are judging (Funder, Kolar, & Blackman, 1995; Kenny, 1994), and friends judge targets more accurately than strangers do (Funder & Colvin, 1997).

Studies continue to investigate whether some people are better judges of personality than others and whether there are identifiable traits that correlate with judges' accuracy. John and Robins (1994) recently found that a narcissistic view of self is associated with lower accuracy in judging others' personalities. Ambady, Hallahan, and Rosenthal (1995) found that low expressiveness, sociability, and self-esteem tend to be associated with greater accuracy in judging others' personality. Some researchers have reported a link between extraversion and accuracy at judging personality and emotions (e.g., Akert and Panter, 1988; Funder & Harris, 1986), while others have not (see Riggio & Friedman, 1982, and Rosenthal, 1979).

In a recent meta-analysis, Davis and Kraus (1997) examined the relationship between a number of individual difference measures and what they termed "empathic ability"—that is, accuracy in judging others' emotions, interpersonal relationships, and personality. General intelligence proved to be the trait most reliably associated with empathic accuracy (mean correlation = .23). Given that many studies have used college students as judges, which creates a restricted range of sampled intelligence, this effect size may in fact underestimate the relationship between intelligence and accuracy (see also, Harris, Vernon, & Jang, 1999). Davis and Kraus reported significant links between empathic accuracy and the traits

of interpersonal trust, reputed social sensitivity, socialization, and responsibility. In terms of the five-factor model, these traits seem most likely to overlap with facets of Agreeableness and Conscientiousness. Interestingly, Davis and Kraus found no relationship between self-report measures of empathy and actual empathic accuracy.

To date, little research has systematically studied possible links between gender-related personality traits (i.e., measures of masculinity and femininity) and accuracy in judging personality. However, findings on gender differences in accuracy provide circumstantial evidence that such links may exist. Many studies indicate, for example, that women are on average more accurate than men in decoding nonverbal expressions of emotions (Hall, 1984), and some studies also report that women outperform men in judging some personality traits (e.g., Ambady, Hallahan, & Rosenthal, 1995). If women are at times more accurate than men, a reasonable inference would seem to be that feminine individuals might similarly be more accurate than masculine individuals.

Studying the relation between gender-related traits and judgmental accuracy is complicated by the fact that masculinity and femininity have been assessed in several different ways (Lippa, 1999; Lippa & Connelly, 1990). The measures most commonly used over the past 25 years are scales of masculine instrumentality and feminine expressiveness (Bem, 1974, 1981; Spence, Helmreich, & Stapp, 1974). Gender diagnosticity (GD) measures, which assess the degree to which individuals' occupational preferences and interests are male- or female-typical, have also been used in recent research on gender-related traits (Lippa & Connelly, 1990; Lippa, 1991, 1995, 1998a, 1998b). Instrumentality and expressiveness overlap substantially with Big Five factors and facets, whereas GD measures do not. Conversely, GD measures overlap with a fundamental dimension of vocational interests—the People-Things dimension—whereas instrumentality and expressiveness do not (Lippa, 1998b, 1999).

Feminine expressiveness would seem a particularly likely candidate to correlate with judgmental accuracy, given that it assesses nurturance and sensitivity to others and empirically correlates with self-report measures of empathy and interpersonal sensitivity (Cook, 1985). GD would also seem a possible correlate of judgmental accuracy, given its overlap with the People-Things dimension, which taps the degree to which individuals prefer activities that deal with people versus activities that deal with mechanical and inanimate things (Lippa, 1998b). Because individuals on the People side of the People-Things dimension are more oriented to managing, instructing, and working with people, it would seem reasonable that such individuals might be more motivated, experienced, and accurate judges of

others. On the other hand, because individuals on the Things side of the People-Things dimension are not oriented to others, it seems similarly reasonable that they might be less motivated, experienced, and accurate judges of others.

The research to be reported here systematically investigated the relationship between judges' personal characteristics (their gender, personality traits, and intelligence) and their accuracy in judging others' personality. Judges in our study were assessed on the Big Five personality traits and three kinds of gender-related traits—masculine instrumentality, feminine expressiveness, and GD. Our predictions, based on previous research, were as follows: Intelligence should positively correlate with judgmental accuracy. To the extent there are gender differences, women should show greater accuracy than men (although such differences are inconsistent in the research literature; see Eisenberg & Lennon, 1983; Graham & Ickes, 1997). For Big Five traits, there is sketchy evidence suggesting that extraversion, agreeableness, and conscientiousness might be linked to accuracy. Davis and Kraus's (1997) review indicated no relationship between neuroticism and empathic accuracy, and there is little existing evidence on the relationship between openness and accuracy or on the relationship between gender-related traits and accuracy. We hypothesized that feminine expressiveness and GD measures were the two gender-related traits most likely to correlate with accuracy in judging others' personalities, with expressive and female-typical individuals likely to display higher levels of accuracy.

In the current study, the personality traits that judges were asked to perceive in others included two Big Five traits—extraversion and neuroticism—as well as a non-Big-Five trait, masculinity-femininity (M-F). We chose extraversion as one of the traits to be judged because considerable research indicates that it is the most observable and accurately judged of the Big Five traits. In some sense, it is the standard against which the judgment of other traits can be compared. Neuroticism was selected because of its central status in most recent personality trait taxonomies (Watson & Clark, 1984; Wiggins & Trapnell, 1997). Furthermore, because neuroticism is likely to be subject to more expressive control than extraversion, the judgment of neuroticism provides an interesting comparison to the judgment of extraversion (see Lippa, 1977, 1983). In addition, the judgment of neuroticism has great relevance to social empathy both in everyday life and in clinical practice (Harrigan & Rosenthal, submitted). Though not judged as accurately as extraversion, neuroticism can nonetheless be judged at better than chance levels of accuracy in "zero acquaintance" settings (Borkenau & Liebler, 1995; Watson, 1989).

M-F was chosen as the third trait to be judged because of its independence from Big Five traits (Lippa, 1991, 1999). More than extraversion and neuroticism, M-F appears to be enacted through conventional, culturally determined nonverbal cues, and this would suggest that judgments of M-F might show higher consensus accuracy (i.e., agreement among judges) than judgments of extraversion or neuroticism (Lippa, 1983, 1998a). Although neglected in recent accuracy research, M-F is a trait that can be accurately judged in “zero acquaintance” settings (Lippa, 1998a).

In the current study, judges (53 college men and 56 college women) were asked to observe brief videotaped segments of 32 male strangers and then rate each of these strangers on the three targeted personality traits of extraversion, neuroticism, and M-F. Two kinds of accuracy were computed: 1) “consensus accuracy,” which measured the correlation between judges’ ratings of targets and corresponding ratings of the same targets made by previous judges and 2) “trait accuracy,” which measured the correlation between participants’ trait ratings of targets and the targets’ assessed personality on the corresponding traits. Using correlations as indices of accuracy rather than difference scores helped us avoid many of the methodological problems that plagued early accuracy research (see Ambady, Hallahan, & Rosenthal, 1995, and Cronbach, 1955). Our accuracy scores assessed the degree to which judges’ ordering of stimulus subjects on a given trait (e.g., extraversion) matched the other judges’ ordering or the ordering provided by stimulus subjects’ self-reported personality. Because correlations are standardized measures, they are not influenced by response sets such as “elevation” (a judge’s overall tendency to rate targets high or low on a trait) or the tendency to use non-extreme or extreme ratings (i.e., to show low or high variance of ratings).

Recent accuracy research suggests that judges show more agreement in inferring personality traits from nonverbal cues (that is, consensus accuracy) than they show accuracy in judging targets’ assessed personality traits from those same nonverbal cues (trait accuracy) (Gifford, 1994; Lippa, 1998a). The current research provides a new test of this finding, specifically for judgments of extraversion, neuroticism, and M-F. We hypothesized that, in general, consensus accuracy would be greater than trait accuracy. It should be noted that in the current study, judges viewed very brief videotapes of targets as they delivered standardized talks (i.e., as they roleplayed being advertisers). Thus, in the current research judgments of personality were based primarily on paralinguistic and nonverbal cues and on physical appearance and grooming.

Because we obtained two different kinds of accuracy measures (consensus accuracy and trait accuracy) for three different judged personality

traits (extraversion, neuroticism, and M-F), we could examine the degree to which various measures of accuracy correlate with one another. To the extent that accuracy measures proved to be strongly intercorrelated, this would argue for the existence of a general ability to judge others' personalities. In contrast, low correlations would instead suggest the existence of specific abilities for judging specific traits. If judges' traits (i.e., gender, personality, intelligence) showed different patterns of correlation with various accuracy measures, this too would offer evidence for the existence of specific abilities.

Method

Participants

Judges were 53 male and 56 female undergraduates at California State University, Fullerton. Thirty-five percent of these judges were Asian, 29 percent Hispanic, 21 percent Caucasian, 5 percent African American, and 2 percent Middle Eastern. Seven percent reported being of mixed ethnicity, and 1 percent did not categorize themselves on ethnicity. The median age was 19.

Video Targets and Criteria of Accuracy

Targets in the current study consisted of 32 college men who had been videotaped as part a previous research project (see Lippa, 1998a, for methodological details). These men had been briefly (on average 30 seconds) videotaped as they role-played being TV announcers advertising eyeglasses. Videotape segments portrayed the targets' full bodies. Because targets used props during their talks (for example, pairs of glasses that they demonstrated), they displayed numerous gestures and movements during their talks. However, because targets had been provided with a suggested script before their talks, the verbal content of talks was relatively standardized. Judged video segments included audio as well as visual information.

As participants in a previous study, video targets had been assessed on a number of personality traits. Specifically, they had completed a Big Five questionnaire that asked them to rate themselves on a number of trait adjectives loading highly on each of the Big Five dimensions. In addition, they had been assessed on several scales of negative affectivity, including the Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), the Rosenberg Self-Esteem Scale (Rosenberg, 1965), and a measure

of interpersonal problems (Horowitz, Rosenberg, Bauer, Ureño, & Villaseñor, 1988). Targets had also been assessed on several GD measures, which assessed how male- or female-typical their preferences were for various occupations, hobbies, and everyday activities.

To compute “trait accuracy” in the current research, we used several criteria of accuracy: 1) targets’ scores on the Big Five extraversion scale served as the accuracy criteria for judgments of their extraversion. 2) targets’ mean scores on four negative affectivity scales served as the accuracy criteria for judgments of neuroticism. Specifically, scores on the following scales were converted to Z-scores and averaged: Big Five neuroticism, Beck depression, self-esteem (reversed), and mean interpersonal problems. Prior research has shown that in other populations of judges, this composite correlated more strongly with targets’ judged neuroticism than did any of the individual component scales (Arad and Lippa, 1999). 3) Targets’ GD scores based on everyday activities served as accuracy criteria for judgments of M-F. These GD scores, which assessed how male- or female-typical a target’s everyday activities were, showed the strongest relationship to judged M-F in previous research (Lippa, 1998a). In the current study, a participant’s “trait accuracy” score was simply the correlation of his or her ratings of all of the 32 targets on a given trait (e.g., extraversion) with the targets’ assessed personality on that same trait.

To compute “consensus accuracy” measures, we used ratings of targets’ extraversion, neuroticism, and M-F obtained in previous research as accuracy criteria (Lippa, 1998a). Specifically, in previous research six research assistants (three men and three women) had rated each of the videotaped targets on extraversion, neuroticism, and M-F. The mean ratings of these six previous judges served as criteria for “consensus accuracy.” The reliabilities (α) of these ratings across the six judges were respectively .85, .67, and .79. Thus, in the current research, a judge’s “consensus accuracy” score was simply the correlation of his or her ratings of all the 32 targets on a given trait with the mean ratings made by the six previous judges of all 32 targets on the same trait. Thus “consensus accuracy” was a measure of how much a judge in the current study agreed with the mean ratings made by six previous judges.

Procedure

In groups ranging from 5 to 12 individuals, judges completed questionnaires and rated video targets. Judges completed a questionnaire packet that assessed demographic information and that included a short measure of the Big Five personality traits (see Lippa, 1991, 1995, for details

about these scales), the Personal Attributes Questionnaire (PAQ; a measure of masculine instrumentality and feminine expressiveness; Spence & Helmreich, 1978; Spence, Helmreich, & Stapp, 1974), and questionnaires that asked respondents to rate their preferences for 70 occupations and 60 hobbies (these were used to compute GD measures).

The Wonderlic Personnel Test (WPT, 1992) was also administered to judges during the experimental session. The Wonderlic is a timed 12-minute group test of general cognitive ability. It consists of 50 multiple-choice and short-answer items including verbal, mathematical, analytical, and pictorial questions. McKelvie (1989) reported split-half reliabilities for the Wonderlic ranging from .88 to .94. Test-retest reliabilities range from .82 to .94, and Wonderlic scores correlate strongly (.91) with IQ as assessed by the Wechsler Adult Intelligence Scale (WPT, 1992, p. 19–21).

After completing the personality questionnaire and the Wonderlic test, participants were shown video segments of the 32 target men and asked to rate each target on the following dimensions: *masculine*, *anxious*, *introverted*, *calm*, *extraverted*, and *feminine*. Ratings were made on a scale that ranged from “1—not at all” to “7—extremely.” To compute trait judgment scores for the three dimensions of personality of interest to us (extraversion, neuroticism, and M-F), we computed three summed ratings: extraversion (the sum of *extraverted* and reversed *introverted*), neuroticism (sum of *anxious* and reversed *calm*), and M-F (sum of *masculine* and reversed *feminine*).

Results

Personality and Intelligence Measures: Computation and Reliability

Big Five scales assessing judges' personalities were computed in standard ways (see Lippa, 1991, 1995). The reliabilities (alpha) of Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness were respectively .76, .67, .73, .78, and .71. PAQ instrumentality and expressiveness scales were also scored in standard ways, and computed reliabilities for these scales were respectively .76 and .70.

GD scores were computed by applying discriminant analyses to sets of occupational and hobby preferences, using sex of participants (i.e., judges) as the grouping variable (for methodological details, see Lippa & Connelly, 1990; Lippa, 1991, 1995, 1998a, 1998b). GD scores are the estimated probability that a judge is male or female based on the judge's pattern of occupational or hobby preferences. In essence, GD scores assess how male- or female-typical a judge's occupational and hobby preferences are.

The reliabilities of GD scores based on occupational preferences were .93 for all judges, .78 for male judges, and .72 for female judges. The reliabilities of GD scores based on hobby preferences were .90 for all judges, .67 for male judges, and .65 for female judges.

The Wonderlic test was scored in a standard fashion (see WPT, 1992). For our sample of 108 judges, the mean Wonderlic score was 23.57 with a standard deviation of 5.76. These values are comparable to normative statistics presented in the Wonderlic manual (WPT, 1992), which indicates that our sample represented a broad range of general intelligence.

Accuracy as a Function of Gender, Type of Accuracy, and Trait Judged

To assess whether accuracy varied as a function of judges' gender, the type of accuracy assessed, and the trait judged, we conducted a $2 \times 2 \times 3$ repeated-measures analysis of variance (ANOVA) on accuracy scores.¹ Gender (male, female) constituted a between-subjects factor, whereas "type of accuracy" (consensus accuracy, trait accuracy) and "trait judged" (extraversion, neuroticism, M-F) constituted within-subjects factors.

The ANOVA showed that the intercept (i.e., mean accuracy) was significantly different from zero, $F(1,105) = 588.91, p < .001$. Thus judges on average displayed a significant degree of accuracy. The ANOVA showed no gender difference in accuracy, $F(1,105) = .001, ns$. However, there was a significant main effect for the type of accuracy, $F(1,105) = 353.20, p < .001$ and also for trait being judged, $F(2,210) = 17.05, p < .001$.

Furthermore, there was a significant interaction between the type of accuracy and the trait being judged, $F(2,210) = 111.10, p < .001$. The cell means illustrating these effects are presented in Table 1. Each cell mean is presented with a 95 percent confidence interval.

As Table 1 shows, the main effect for "type of accuracy" resulted from the fact that consensus accuracy was considerably higher than trait accuracy (.45 versus .21). The main effect for "trait judged" reflected the fact that mean accuracy for judgments of extraversion and M-F (.35 and .41, respectively) were higher than accuracy for judgments of neuroticism (.24). Finally, the interaction between "type of accuracy" and "trait judged" resulted from the fact that the difference between consensus accuracy and trait accuracy was particularly large for ratings of M-F (consensus accuracy = .64 vs. trait accuracy = .18). The high level of consensus accuracy for judgments of M-F indicates that current judges agreed strongly with past judges about which targets were masculine and which were not.

Paired-data *t*-tests allowed us to compare means from pairs of cells in Table 1. Three paired-data *t*-tests compared consensus accuracy and trait

TABLE 1

Judges' Mean Accuracy by Type of Accuracy and Judged Trait

Type of accuracy	Judged Trait			Row means
	Masculinity-femininity	Extraversion	Neuroticism	
Consensus Accuracy	.64 ± .02	.42 ± .06	.30 ± .06	.45
Trait Accuracy	.18 ± .03	.28 ± .04	.17 ± .04	.21
Column Means	.41	.35	.24	

Note. N = 107.

Two participants were not included because of missing values.

Cell means are accompanied by 95 percent confidence intervals.

accuracy, separately for each of the three judged traits. All of these comparisons were significant ($p < .001$), indicating that consensus accuracy exceeded trait accuracy for each of the judged traits. Paired-data t -tests also compared consensus accuracy for different pairs of traits (e.g., consensus accuracy for M-F vs. consensus accuracy for extraversion) and also trait accuracy for different pairs of traits (e.g., trait accuracy for M-F vs. trait accuracy for extraversion). All of these comparisons were significant ($p < .01$), except for the comparison of trait accuracy for M-F (.18) and trait accuracy for neuroticism (.17). Thus, these tests showed that for *consensus measures* of accuracy, M-F was judged more accurately than extraversion, which in turn was judged more accurately than neuroticism. However, for *trait measures* of accuracy, extraversion was judged more accurately than either M-F or neuroticism, which were both judged with the same level of accuracy.

Relation Between Judges' Personality, Intelligence, and Accuracy

Table 2 presents correlations between judges' assessed traits (i.e., personality and intelligence) and their accuracy in judging targets. We additionally computed an overall accuracy score, which was simply the mean of the Z-transforms of the six accuracy scores described earlier. The reliability (alpha) of this composite was .61.

Judges' intelligence correlated significantly with consensus accuracy in judging extraversion, trait accuracy in judging extraversion, consensus accuracy in judging M-F, and overall accuracy. Judges' openness correlated

TABLE 2

Correlation of Judges' Personality and Intelligence with Accuracy Scores

Judges' traits	Consensus accuracy for:			Trait accuracy for			Overall accuracy
	Masculinity-femininity	Extraversion	Neuroticism	Masculinity-femininity	Extraversion	Neuroticism	
Extraversion	.09	.02	-.20*	.10	.01	-.04	-.04
Agreeableness	.14	.02	.02	.01	.17	.09	.11
Conscientiousness	.17	.05	.04	.05	.08	.05	.10
Neuroticism	.10	.00	-.02	-.07	-.07	-.06	-.02
Openness	-.09	.11	-.36**	-.10	.02	-.30**	-.20*
PAQ-M	-.01	.05	-.12	.00	-.02	-.10	-.07
PAQ-F	.18	.13	.01	-.01	.16	.10	.16
Intelligence	.26**	.41***	.14	.13	.34**	-.05	.36**

Note. N = 107 to 109. N's varied because of missing values in personality scores.

*Two-tailed $p < .05$. **Two-tailed $p < .01$. ***Two-tailed $p < .001$

negatively with consensus accuracy in judging neuroticism, trait accuracy in judging neuroticism, and with overall accuracy. Correlations computed separately for men and women showed much the same pattern, although for the sake of brevity, they are not presented here.

Except for openness, personality measures generally showed little relationship to accuracy scores. Correlations for GD scores are not shown in Table 2 because GD correlates strongly with sex of judge for men and women combined. When GD scores were correlated with accuracy measures for men and women separately, no correlations were significant.

The Relation of Gender to Specific Accuracy Measures

Although the ANOVA reported earlier found no evidence for a gender difference in overall accuracy, we also examined possible links between gender and individual accuracy scores. That is, we performed *t*-tests that compared men's and women's means for each of the six accuracy scores. Only one of these tests yielded a significant result: Women tended to show higher trait accuracy when judging neuroticism than men did ($t(107) = -1.70$, 1-tailed $p < .05$; mean accuracy = .14 for men, and .20 for women).

Intercorrelations of Accuracy Scores

Table 3 shows the intercorrelations of the six accuracy scores. The highest correlations were between consensus and trait accuracy scores for the same trait. However, these relatively high correlations were probably due largely to a method artifact, which resulted from the fact that consensus and trait accuracy correlations for a given trait were computed with a common data vector (e.g., the judge's ratings of the 32 targets on a given trait). Although generally positive, the remaining correlations in Table 3 are quite modest in magnitude and do not provide much evidence for a general trait of judgmental accuracy. In general, measures of consensus accuracy show a bit more coherence than do measures of trait accuracy.

Discussion

Like previous researchers, we found that judges showed a significant degree of accuracy when judging strangers' personality from "thin slices" of expressive behavior in a "zero acquaintance" paradigm. In our study, the mean level of consensus accuracy was .45 and the mean level of trait

TABLE 3

Interrcorrelations of Accuracy Scores

	Consensus accuracy			Trait accuracy		
	M-F	Extraversion	Neuroticism	M-F	Extraversion	Neuroticism
Consensus Accuracy M-F		.24*	.22*	.54***	.17	.06
Consensus Accuracy Extraversion			.28**	.00	.76***	-.02
Consensus Accuracy Neuroticism				.17	.27**	.63***
Trait Accuracy M-F					.00	.12
Trait Accuracy Extraversion						.04

*Two-tailed $p < .05$
 **Two-tailed $p < .01$
 ***Two-tailed $p < .001$

accuracy was .21. These are not trivial levels, given that 1) the correlations used to assess accuracy were uncorrected for attenuation due to unreliability of ratings and personality criteria, and 2) the target information judges were exposed to was quite brief and limited.

Contrary to expectation, we did not find an overall gender difference in accuracy, nor did we find that gender-related traits correlated with accuracy, either within the sexes or for the sexes combined. In general, Big Five traits showed little relationship to judgmental accuracy. The one exception was openness to experience, which correlated negatively with accuracy for judgments of neuroticism. Consistent with Davis and Kraus's (1997) meta-analysis, we found that general intelligence proved to be the strongest individual difference correlate of judgmental accuracy. Intelligent people tended

to be better judges of personality, particularly when judging extraversion and to a lesser degree when judging M-F.

Also consistent with previous research (see Davis & Kraus, 1997), the current data provide little evidence for the existence of a general ability to judge others' personalities. Rather, they suggest that the abilities that feed into judgmental accuracy may be quite trait-specific. Participants' accuracy at judging one trait bore little relationship to their accuracy at judging another trait. Furthermore, the individual difference correlates of judgmental accuracy varied depending on the trait being judged. Intelligence was particularly linked to accuracy at judging extraversion and M-F, whereas openness was particularly linked (albeit negatively) to accuracy in judging neuroticism. Similarly, gender was related to trait accuracy in judgments of neuroticism, but not to accuracy in judgments of extraversion or M-F.

In general, our findings hint that the judgment of neuroticism may be different from the judgment of extraversion and M-F. As suggested earlier, the display of neuroticism may be more subject to expressive control and inhibition than the display of extraversion or M-F are (Lippa, 1977, 1983). Our finding that trait accuracy was lowest for judgments of neuroticism seems consistent with this hypothesis.

Because neuroticism is a uniquely emotional kind of trait, it seems likely that the judgment of neuroticism, more than the judgment of extraversion or M-F, is related to the judgment of emotion. Thus, the cognitive processes used to judge neuroticism may be somewhat different from the processes used to judge extraversion and M-F. The finding that intelligence correlated with judges' accuracy in judging extraversion and M-F but not neuroticism seems consistent with this hypothesis. Perhaps the judgment of extraversion is more a matter of conscious information processing (e.g., noticing valid cues of extraversion and integrating them), whereas the judgment of neuroticism may have more in common with unconscious and relatively automatic perceptions of emotion (see Hodges & Wegner, 1997). It is intriguing to note that trait accuracy in judging neuroticism showed the strongest evidence for a gender difference in our study, which is consistent with the finding that the most robust gender differences in accuracy tend to be for judgments of emotion (Hall, 1984).

Our data replicated previous findings that consensus accuracy tends to be greater than trait accuracy, when based on limited nonverbal cues (Gifford, 1994). Furthermore, consensus accuracy proved to be particularly high for judgments of M-F. Lippa (1998a) has speculated that M-F cues, unlike extraversion and neuroticism cues, may reflect cultural conventions (for example conventions about the social meanings of dress, grooming, and mannerisms). In this sense, the display of M-F may be more socially

constructed than the display of extraversion and neuroticism, and this may help account for the especially high consensus accuracy we found for judgments of M-F.

The current data replicated previous findings that extraversion is the most accurately judged of all traits, when accuracy is assessed in terms of the correspondence between judges' ratings and targets' assessed personality. Despite the high "judgability" of extraversion, it is again worth noting again that consensus accuracy was higher for M-F than for extraversion. Because of its unique pattern of high consensus accuracy and relatively low trait accuracy, M-F would seem to be a trait worth including in future studies on the accuracy of personality judgments.

The current findings have a number of implications for future research on individual difference correlates of the accuracy of personality judgments. First, consistent with Davis and Kraus's (1997) meta-analysis, we found that general intelligence proved to be the strongest individual difference predictor of judgmental accuracy, and we therefore recommend that future research on this topic include measures of intelligence as a matter of course. We suspect that intelligence will prove to correlate even more strongly with judgmental accuracy in studies that ask participants to judge personality from complex, extended information, rather than from "thin slices" of relatively impoverished video information (see Ambady & Rosenthal, 1992; Harris, Vernon, & Jang, 1998). As Gordon Allport (1937) observed more than 60 years ago, "Understanding people is largely a matter of perceiving relations between past and present activities, between expressive behavior and inner traits, between cause and effect, and intelligence is the ability to perceive just such relations as these" (p. 514).

The current research also suggests that the judgment of one personality trait may entail different kinds of cognitive processes from the judgment of another. Judgments of neuroticism—particularly when based upon "thin slices" of expressive behavior—may share more in common with the judgment of emotions than do judgments of extraversion or M-F. This distinction between judgments of neuroticism and judgments of other traits may have implications for which individual difference variables will correlate with judgmental accuracy. The current findings suggest that intelligence is more likely to correlate with accuracy in judging non-emotional kinds of traits, whereas openness and gender are more likely to be linked to accuracy in judging emotional kinds of traits.

Our study found no evidence that judges' gender-related traits—that is, their degree of masculinity and femininity—are related to judgmental accuracy. Although no single study can conclusively demonstrate a null finding, the current results lessen the likelihood that gender-related traits

will prove to be important correlates of judgmental accuracy in “zero acquaintance” situations. Once again, however, it is important to note that the current study focused on judgments of strangers’ personality from “thin slices” of expressive behavior. It is possible that gender and gender-related traits might correlate more strongly with accuracy in real-life and long-term kinds of judgment settings. Furthermore, gender and gender-related traits might be linked to individuals’ *motivation* to judge others’ personality in real-life settings (see Graham & Ickes, 1997).

The current data showed intriguing, if unexplained, relationships between openness and judgmental accuracy. Openness was negatively correlated with accuracy in judging neuroticism. Ironically, openness has been the least investigated Big Five trait in relation to judgmental accuracy. The current findings suggest that this neglect may be unwarranted. Openness is the Big Five dimension most linked to actual and self-reported intelligence (Loehlin, McCrae, Costa, & John, 1998; McCrae & Costa, 1997). However, in our study openness was related to different kinds of judgmental accuracy than intelligence was, and furthermore, the direction of the relationship was opposite for these two traits—intelligence was positively related to accuracy, whereas openness was negatively related to accuracy. Given that people who are high on openness are likely to be thoughtful and thus engage in more complex, conscious cognitive processing of information, the current findings suggest that too much thought may at times interfere with gut-level judgments of emotional traits (see Wilson, Dunn, Kraft, & Lisle, 1989, for related research on how introspection may sometimes interfere with the self-perception of attitudes). Clearly, additional research is warranted that investigates the relationship between Big Five traits, intelligence, and accuracy at judging others’ personality.

Note

1. Because accuracy scores were correlation coefficients, we also conducted analyses on Z-transforms of scores. The results for these analyses were very similar to the results we report for non-transformed scores. Because raw correlations are intuitively understood more readily than Z-transforms, we report the results here in terms of raw correlations.

References

- Akert, R. M. & Panter, A. T. (1988). Extraversion and the ability to decode nonverbal communication. *Personality and Individual Differences*, 9, 965–972.
- Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology*, 55, 387–395.

- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt, Rinehart, & Winston.
- Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology*, 69, 518–529.
- Ambady, N. & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 111, 256–274.
- Arad, S. & Lippa, R. (1999). Can trait anxiety be accurately judged from brief video segments? Poster presented at the annual meeting of the American Psychological Society, Denver, CO.
- Beck, A. T, Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 53–63.
- Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, 42, 165–172.
- Bem, S. L. (1981). *Bem Sex Role Inventory professional manual*. Palo Alto, CA: Consulting Psychologist Press.
- Borkenau, P. & Liebler, A. (1993). Convergence of stranger ratings of personality and intelligence with self-ratings, partner-ratings, and measured intelligence. *Journal of Personality and Social Psychology*, 65, 546–553.
- Borkenau, P. & Liebler, A. (1995). Observable attributes as manifestations of personality and intelligence. *Journal of Personality*, 63, 1–25.
- Colvin, C. R. (1993). “Judgable” people: Personality, behavior, and competing explanations. *Journal of Personality and Social Psychology*, 64, 861–873.
- Cook, E. P. (1985). *Psychological androgyny*. New York: Pergamon Press.
- Cronbach, L. J. (1955). Processes affecting scores on “understanding of others” and “assumed similarity.” *Psychological Bulletin*, 52, 177–193.
- Davis, H. M. & Kraus, A. L. (1997). Personality and empathic accuracy. In W. Ickes (Ed.), *Empathic accuracy* (pp. 145–165). New York: The Guilford Press.
- Eisenberg, N. & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin*, 94, 100–131.
- Funder, D. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102, 652–670.
- Funder, D. C. & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology*, 55, 149–158.
- Funder, D. C. & Colvin, C. R. (1997). Congruence of others’ and self-judgments of personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality* (pp. 617–647). San Diego, CA: Academic Press.
- Funder, D. C. & Dobroth, K. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology*, 52, 409–418.
- Funder, D. C. & Harris, M. J. (1986). On the several facets of personality assessment: The case of social acuity. *Journal of Personality*, 54, 528–550.
- Funder, D. C., Kolar, D. C., & Blackman, M. C. (1995). Agreement among judges of personality: Interpersonal relations, similarity, acquaintanceship. *Journal of Personality and Social Psychology*, 69, 656–672.
- Funder, D. C. & West, S. G. (1993). Consensus, self-other agreement, and accuracy in personality judgment: An introduction. *Journal of Personality*, 61, 457–475.
- Gifford, R. (1994). A lens-mapping framework for understanding the encoding and decoding of interpersonal dispositions in nonverbal behavior. *Journal of Personality and Social Psychology*, 66, 398–412.
- Graham, T. & Ickes, W. (1997). When women’s intuition isn’t greater than men’s. In W. Ickes (Ed.), *Empathic accuracy* (pp. 145–165). New York: The Guilford Press.
- Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive styles*. Baltimore: Johns Hopkins University Press.

- Harrigan, J. A. & Rosenthal, R. (submitted). Detecting state and trait anxiety from auditory and visual cues: A meta-analysis.
- Harris, J. A., Vernon, P. A., & Jang, K. L. (1998). Intelligence and personality characteristics associated with accuracy in rating a co-twin's personality. *Personality and Individual Differences*, 26, 85–97.
- Hodges, S. D. & Wegner, D. M. (1997). Automatic and controlled empathy. In W. Ickes (Ed.), *Empathic accuracy* (pp. 311–339). New York: The Guilford Press.
- Horowitz, L. M., Rosenberg, S. E., Bauer, B. A., Ureño, G., & Villaseñor, V. S. (1988). Inventory of Interpersonal Problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology*, 56, 885–892.
- John, O. P. & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and narcissism. *Journal of Personality and Social Psychology*, 66, 206–219.
- Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review*, 98, 155–163.
- Kenny, D. A. (1994). *Interpersonal perception*. New York: Guilford Press.
- Kenny, D. A. & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin*, 102, 390–402.
- Kenny, D. A., Horner, C., Kashy, D. A., & Chu, L. (1992). Consensus at zero acquaintance: Replication, behavioral cues, and stability. *Journal of Personality and Social Psychology*, 62, 88–97.
- Levesque, M. J. & Kenny, D. A. (1993). Accuracy of behavioral predictions at zero acquaintance: A social relations analysis. *Journal of Personality and Social Psychology*, 65, 1178–1187.
- Lippa, R. (1977). Expressive control, expressive consistency, and the correspondence between expressive behavior and personality. *Journal of Personality*, 46, 438–461.
- Lippa, R. (1983). Expressive behavior. In L. Wheeler & P. Shaver (Eds.), *Review of personality and social psychology* (Vol. 4). Beverly Hills: Sage.
- Lippa, R. (1991). Some psychometric characteristics of gender diagnosticity measures: reliability, validity, consistency across domains, and relationship to the Big Five. *Journal of Personality and Social Psychology*, 61, 1000–1011.
- Lippa, R. A. (1995). Gender-related individual differences and psychological adjustment in terms of the Big Five and Circumplex models. *Journal of Personality and Social Psychology*, 6, 1184–1202.
- Lippa, R. A. (1998a). The nonverbal display and judgment of extraversion, masculinity, femininity, and gender diagnosticity: A lens model analysis. *Journal of Research in Personality*, 32, 80–107.
- Lippa, R. (1998b). Gender-related individual differences and the structure of vocational interests: The importance of the "People-Things" dimension. *Journal of Personality and Social Psychology*, 74, 996–1009.
- Lippa, R. A. (1999). On deconstructing and reconstructing masculinity-femininity. Unpublished manuscript. California State University, Fullerton.
- Lippa, R. A. & Connelly, S. C. (1990). Gender diagnosticity: A new Bayesian approach to gender-related individual differences. *Journal of Personality and Social Psychology*, 59, 1051–1065.
- Loehlin, J. C., McCrae, R. R., Costa, P. T., Jr., & John, O. P. (1998). Heritabilities of common and measure-specific components of the Big Five personality factors. *Journal of Research in Personality*, 32, 431–453.
- McCrae, R. R. & Costa, P. T. (1997). Conceptions and correlates of openness to experience. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality* (pp. 825–847). San Diego, CA: Academic Press.
- McKelvie, S. J. (1989). The Wonderlic Personnel Test: Reliability and validity in an academic setting. *Psychological Reports*, 65, 161–162.
- Park, B. & Judd, C. M. (1989). Agreement on initial impressions: Differences due to per-

- ceivers, trait dimensions and target behaviors. *Journal of Personality and Social Psychology*, 56, 493–505.
- Riggio, R. E. & Friedman, H. S. (1982). The interrelationship of self-monitoring factors, personality traits, and nonverbal social skills. *Journal of Nonverbal Behavior*, 7, 33–45.
- Rosenberg, M. (1965). *Society and adolescent self-image*. Princeton, NJ: Princeton University Press.
- Rosenthal, R. (1979). *Skill in nonverbal communication: Individual differences*. Cambridge, MA: Oelgeschlager, Gunn, Hain.
- Spence, J. T. & Helmreich, R. L. (1978). *Masculinity and femininity: Their psychological dimensions, correlates, and antecedents*. Austin: University of Texas Press.
- Spence, J. T., Helmreich, R. L., & Stapp, J. (1974). The personal attributes questionnaire: A measure of sex role stereotypes and masculinity-femininity. *JSAS, Catalog of Selected Documents in Psychology*, 4, 43–44.
- Taft, R. (1955). Accuracy of empathic judgments of acquaintances and strangers. *Journal of Personality and Social Psychology*, 3, 600–604.
- Watson, D. (1989). Strangers' ratings of five robust personality factors: Evidence of a surprising convergence with self-reports. *Journal of Personality and Social Psychology*, 57, 120–128.
- Watson, D. & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin*, 96, 465–490.
- Wiggins, J. S. & Trapnell, P. D. (1997). *Personality structure: The return of the Big Five*. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality* (pp. 737–765). San Diego, CA: Academic Press.
- Wilson, T. D., Dunn, D. S., Kraft, D., & Lisle, D. J. (1989). Introspection, attitude change, and attitude-behavior consistency: The disruptive effects of explaining why we feel the way we do. In L. Berkowitz (Ed.), *Advance in experimental social psychology* (Vol. 22). San Diego: Academic Press.
- WPT, *Wonderlic personnel test & scholastic level exam: User's manual*. (1992). Libertyville, IL: Wonderlic Personnel Test, Inc.